SOUND VARIATIONS BY RECURRENT NEURAL NETWORK SYNTHESIS

Ken'ichi Ohya
Dept. of Electronics and Computer Science,
Nagano National College of Technology,
716 Tokuma, Nagano City, Nagano 381-8550, JAPAN
E-mail: ohya@ei.nagano-nct.ac.jp

ABSTRACT: A new sound synthesis technique, recurrent neural network sound synthesis was shown in 1995. Recurrent neural network is a neural network that each neuron connects recurrently.

In this paper, sound variations by this recurrent neural network model are shown. One variation is from relatively small network, and another variation is from relatively large network. The former is suitable for realtime sound synthesis, and the latter is for modifying the sound after resynthesis of the original sound.

1 Introduction

Study of sound synthesis has a long history in computer music and many models have been presented.

A new sound synthesis model using recurrent neural network was shown in 1995, and its capability of learning nonlinear dynamics was presented [Ohya, 1995].

In this paper, sound variations using recurrent neural network model are shown. One sound variation is from relatively small network, and another sound variation is from large network.

Recurrent neural network is a neural network that each neuron connects recurrently. RNNS (Recurrent Neural Network Synthesis) belongs to nonlinear sound synthesis. The following is some features of RNNS: 1) dynamics of neurons are directly used for waveform itself 2) RNNS is like FM synthesis, each neuron corresponds to each operator in FM synthesis model 3) complex waveforms are produced from relatively small number of neurons 4) resynthesis is also possible with relatively large number of neurons by learning algorithm.

2 Single Neuron Model and Recurrent Neural Network

As a single neuron model, a continuous-time, continuous-variable neuron model is adopted. Therefore output value from any single neuron can be directly used for waveform.

Equation of dynamics of each neuron in a recurrent neural network is given ([Ohya, 1995]) as

$$\tau_i \frac{du_i}{dt} = -u_i + f(\sum_{j=1}^{n} W_{ij}u_j) + I_i$$

where $u_i(t)$ is the i-th unit output at a time t, τ_i a time delay constant, $f(x)$ a sigmoid function, I_i an external input of the i-th unit, W_{ij} a connection weight from the j-th unit to the i-th unit.

3 Sound Variation 1: Small Network of Several Neurons

Recurrent Neural Network can generate very complex dynamics pattern because of its recurrent connections even if it is composed of relatively small number of neurons.

As the 1st variation of RNNS, sound synthesis by small recurrent neural network is shown. This model is useful for realtime software sound synthesis because its relatively less computation.

3.1 1 pair model

"1 pair model" is composed of one pair neuron and another output neuron (*Fig. 2*). Each neuron of the pair neuron is connected each other and is also connected to itself. This pair neuron is the smallest recurrent neural network architecture. The output neuron receives two output values from the pair neuron and computes output value, which is waveform itself.

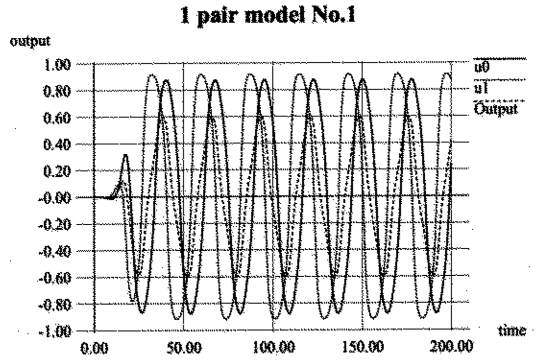

Figure 1: waveform example of 1 pair model

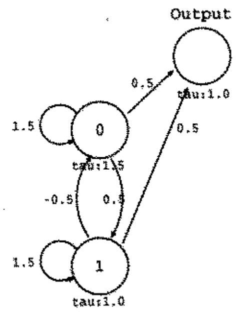

Figure 2: 1 pair model architecture and a parameter example

Dynamics of the pair neuron is completely described by eight parameters, 2 initial values, 2 time delay constants, 4 connection weight values. In an 8-dimension space of parameters, the region which gives this system lasting oscillation seems to be small. One of such region is given if weight values connecting each other are asymmetric. In this condition, 2 output values of a pair neuron spikes by turns, and this pair neuron functions as a kind of oscillator. It is also possible to use only one of the outputs from the pair neuron as waveform itself, but the third neuron, the output neuron, is used for producing more complex, rich waveform.

An example of a waveform of 1 pair model is shown in *Fig. 1*, where "u0" is the output from the neuron No.0, "u1" is the output from the neuron No.1, "Output" is the output from the output neuron, that is waveform itself.

Frequency of the waveform mainly depends on the time delay constant τ, and the total waveform chiefly depends on values of connection weight.

3.2 2 pair model, 3 pair model

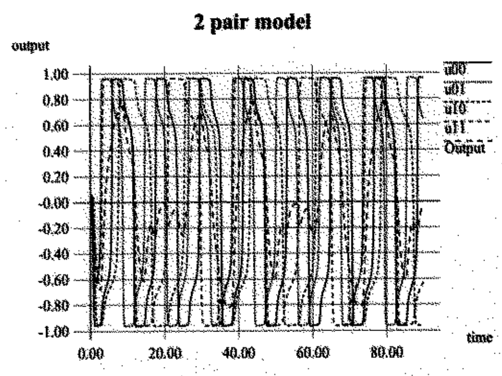

Figure 3: waveform example of 2 pair model

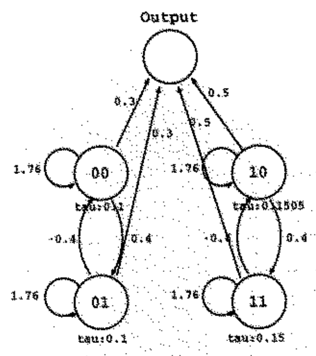

Figure 4: 2 pair model architecture and a parameter example

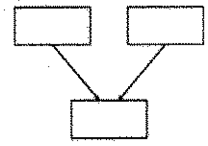

Figure 5: 3 operator model in FM

2 pair model is composed of 2 pair neurons and 1 output neuron (*Fig. 4*). This model has 2 oscillators, frequency of each oscillator depends on its time delay constant τ. Therefore this model resembles 3 operator

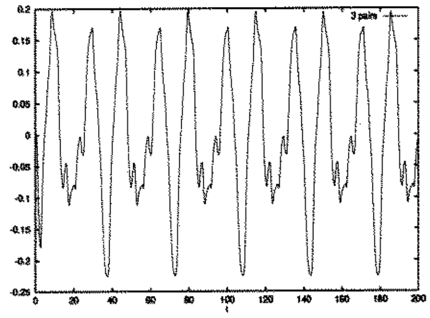

Figure 6: waveform example of 3 pair model

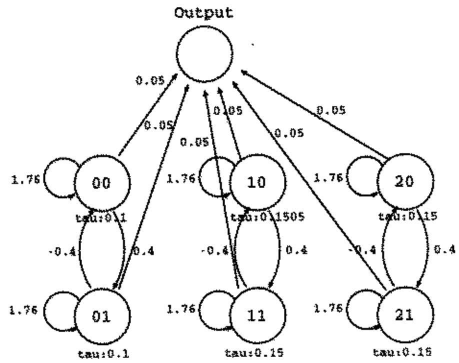

Figure 7: 3 pair model architecture and a parameter example

model in FM synthesis in a sense (*Fig. 5*). Modifying two time delay constants, various waveform can be produced (*Fig. 3*).

As shown in *Fig. 4*, in this example, the ratio of the 2 time delay constants of 2 pair neurons is 1.5, thus harmonics of the sound can be controlled.

And more, by slightly modifying the time delay constant of one of neurons (ex. Neuron No.10 in *Fig. 4*), the output of the right pair neuron can be slightly fluctuated, and the produced sound also can be slightly fluctuated.

In the same way, we can produce more complex waveform by adding more pair neurons. *Fig. 6* is an example of 3 pair model.

4 Sound Variation 2: Resynthesis by Recurrent Neural Network

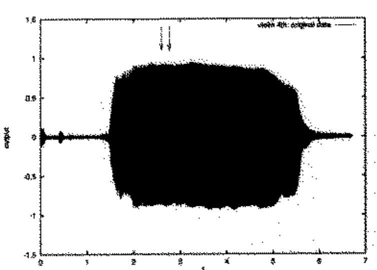

Figure 8: sampled violin sound data

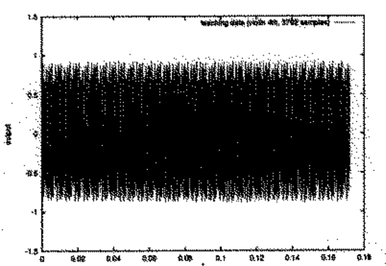

Figure 9: teacher signals

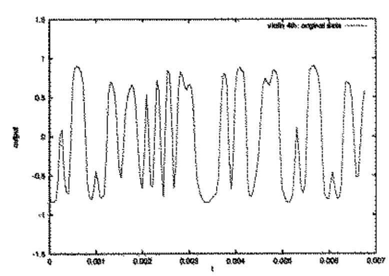

Figure 10: head 300 samples of the teacher signals

Resynthesis of the original sound of the piano was shown by the author ([Ohya, 1995]). But in this case, the amplitude of the piano sound data was always decreasing, and it was rather difficult to select good teacher signals. This time, violin is adopted as a lasting sound source.

The data were sampled in 16-bit integer format at a sampling rate of 22.05 kHz. The note was G, open 4th string. This part, shown between the two arrows in *Fig.8*, is used for teacher signals. It has 3792 samples, 34 periods of 0.17 second. *Fig.10* is the head part of this teacher signals.

To look into the dynamics of the data, 3-dimensional phase space trajectory is also shown (*Fig.11*). Time lag constant is set to 30 samples.

As an architecture of recurrent neural network, an APOLONN is adopted [Murakami, 1991] [Sato, 1990a] [Sato, 1990b] [Sato, 1990c].

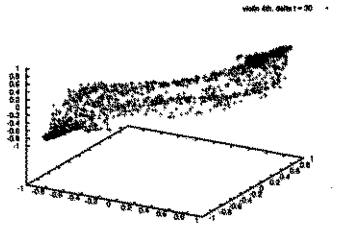

Figure 11: phase trajectory of teacher signals

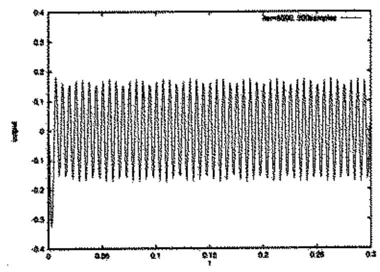

Figure 12: resynthesized waveform: iter=5000, 300 samples

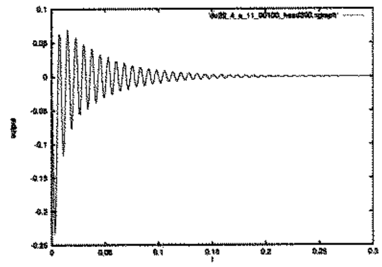

Figure 13: resynthesized waveform: iter=100, 300 samples

The learning process was done by a software simulation on a Workstation Sun Ultra1. 25 pairs of oscillators were used in the simulation. The ratio of the τ_i between two neighboring pairs was 0.9.

One of the leading feature of RNNS is the possibility of resynthesis of the sound using connection weight values. This system is fully described by many differential equations, time unlimited sound resynthesis is possible.

By reading connection weight after learning of 5000 iterations, new sound is resynthesized. The head 300 samples of the sound is shown in *Fig.12*.

On the other hand, *Fig.13* is a waveform of another sound by reading connection weight during learning of only 100 iterations. This figure shows possibilities of a harsh decreasing sound resynthesis without any envelope filters.

5 Summary

Sound variations by RNNS is shown. It is shown that it is possible to synthesis new sound by recurrent neural network. Since this model is composed of relatively small number of neurons, it is possible to synthesis in software.

And another variation of resynthesis by RNNS is shown. Using connection weight values after learning process or during learning process, both resynthesis of the original sound and synthesis of new sound are shown.

References

[Murakami, 1991] Murakami, Y., and Sato, M., (1991). "A recurrent network which learns chaotic dynamics ". *Proc. of ACNN'91*, pp.1–4.

[Ohya, 1995] Ohya, Ken'ichi. (1995). "A Sound Synthesis by Recurrent Neural Network". *Proceedings of the 1995 International Computer Music Conference*, pp.420–423.

[Pearlmutter, 1989] Pearlmutter, B. A. (1989). "Learning state space trajectories in the recurrent neural network". *Neural Computation*, **1** (2), pp.263–269.

[Sato, 1990a] Sato, M. (1990). "A learning algorithm to teach spatiotemporal patterns to recurrent neural networks". *Biological Cybernetics*, **62**, pp.259–263.

[Sato, 1990b] Sato, M., Murakami, Y., & Joe K. (1990). "Chaotic dynamics by recurrent neural networks". *Proceedings of the International Conference on Fuzzy Logic and Neural Networks*, pp.601–604.

[Sato, 1990c] Sato, M., Joe, K., & Hirahara T. (1990). "APOLONN brings us to the real world: Learning nonlinear dynamics and fluctuations in nature". *Proceedings of the International Joint Conference on Neural Networks*, San Diego, **I**, pp.581–587.